p.1

Conditional distribution

We have the joint distribution

$$p(X_1, X_2) = \mathcal{N}\left(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$$

x_1, x_2 – D-dimensional vectors

From the joint, we want to find the conditional $p(x_1, x_2) = \boxed{p(x_1 | x_2)}\, p(x_2)$

$$p(x_2) = \mathcal{N}(\mu_2, \Sigma_{22}) \propto e^{-\frac{1}{2}(x_2 - \mu_2)^T \Sigma_{22}^{-1} (x_2 - \mu_2)}$$

- We want to factor this out from the exponent of the joint distribution, since then we will get the remaining part corresponding to the conditional –

  – $p(x_1, x_2) = p(x_1 | x_2)\, p(x_2)$

- Since the covariance matrix gets inversed in the exponentials (Gaussian form), we would ideally like to rewrite it in a way that it factorises to a block diagonal (then the inverse will operate on the diagonal blocks independently)

  → How do we invert the cov. matrix to keep Σ_{22} isolated?

To do that we resort to linear algebra and identify Schur complement

$$M = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} E+AG & F+AH \\ G & H \end{bmatrix}$$

We want $F + AH = 0 \Rightarrow \underline{\underline{A = -FH^{-1}}}$

So, what we have is

$$\begin{bmatrix} E-FH^{-1}G & 0 \\ G & H \end{bmatrix} \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} = \begin{bmatrix} E-FH^{-1}G & 0 \\ G+HB & H \end{bmatrix}$$

We want $G + HB = 0 \Rightarrow B = -H^{-1}G$

$$\underbrace{\begin{bmatrix} I & -FH^{-1} \\ 0 & I \end{bmatrix}}_{\cancel{A}\ M_A = \begin{bmatrix} I & A \\ 0 & I \end{bmatrix}} \underbrace{\begin{bmatrix} E & F \\ G & H \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} I & 0 \\ -H^{-1}G & I \end{bmatrix}}_{\cancel{B}\ M_B = \begin{bmatrix} I & 0 \\ B & I \end{bmatrix}} = \underbrace{\begin{bmatrix} E-FH^{-1}G & 0 \\ \underbrace{G+HB}_{0} & H \end{bmatrix}}_{C}$$

$$M_A \cdot M \cdot M_B = C$$

p. 3

Let's inverse $M_A M M_B = C$

$$C^{-1} = M_B^{-1} M^{-1} M_A^{-1}$$

$$\Downarrow$$

$$M^{-1} = M_B C^{-1} M_A$$

So, inverting M implies only inverting C. Since C is a nice block diagonal with an isolated H:

$$C = \begin{bmatrix} E - FH^{-1}G & 0 \\ 0 & H \end{bmatrix},$$

we can invert more easily and keep H block nicely isolated (factorised).

We say that $E - FH^{-1}G$ is Schur complement of ~~H in~~ matrix M ($M = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$) w.r.t. H and use a symbol M/H.

Now, let's come back to the exponent of the joint distribution $p(x_1, x_2)$:

$$-\frac{1}{2} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} = \left\{ \text{using } \Sigma/\Sigma_{22} \right\} =$$

$$= -\frac{1}{2} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T \underbrace{\begin{bmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{bmatrix}}_{M_B} \underbrace{\begin{bmatrix} (\Sigma/\Sigma_{22})^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix}}_{C^{-1}} \underbrace{\begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix}}_{M_A} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

We continue with the exponent of $p(x_1, x_2)$:

$$-\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}^T \begin{bmatrix} I & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I\end{bmatrix}\begin{bmatrix} (\Sigma/\Sigma_{22})^{-1} & -(\Sigma/\Sigma_{22})^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & \Sigma_{22}^{-1}\end{bmatrix}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix} =$$

$$= -\frac{1}{2}\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix}^T \left[\begin{array}{c|c} (\Sigma/\Sigma_{22})^{-1} & -(\Sigma/\Sigma_{22})^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ \hline -\Sigma_{22}^{-1}\Sigma_{21}(\Sigma/\Sigma_{22})^{-1} & \Sigma_{22}^{-1}\Sigma_{21}(\Sigma/\Sigma_{22})^{-1}\Sigma_{12}\Sigma_{22}^{-1}+\Sigma_{22}^{-1}\end{array}\right]\begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2\end{bmatrix} =$$

$$= \underbrace{-\frac{1}{2}\left(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)\right)^T(\Sigma/\Sigma_{22})^{-1}\left(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)\right)}_{\mathrm{I}} \underbrace{-\frac{1}{2}(x_2-\mu_2)^T\Sigma_{22}^{-1}(x_2-\mu_2)}_{\mathrm{II}}$$

So, we have 2 terms
and the second one (II) corresponds to $p(x_2)$.
The first one (I) then corresponds to $p(x_1|x_2)$:

$$\boxed{p(x_1|x_2) \propto e^{-\frac{1}{2}\left(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)\right)^T(\Sigma/\Sigma_{22})^{-1}\left(x_1-\mu_1-\Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2)\right)}}$$

Now, let's figure out what the variance and the mean are for this Gaussian distribution :

MEAN : $$\mu_{x_1|x_2} = \mu_1 + \Sigma_{21}\Sigma_{22}^{-1}(x_2-\mu_2)$$

VARIANCE : (covariance for multidimensional x_1 and x_2) $$\Sigma_{x_1|x_2} = \Sigma/\Sigma_{22} = \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

p. 5

So, we have for the conditional $x_1|x_2$:

$$p(x_1|x_2) = \mathcal{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2-\mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

Now, let's see what happens if

(1) x_1 and x_2 are independent

$$\Sigma_{12} = \Sigma_{21} = 0 \Longrightarrow \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = 0 \Longrightarrow \boxed{\Sigma_{x_1|x_2} = \Sigma_{x_1}}$$

$$\Sigma_{12}\Sigma_{22} = 0 \Longrightarrow \boxed{\mu_{x_1|x_2} = \mu_{x_1}}$$

So, $p(x_1|x_2) = p(x_1)$, which makes sense!

(2) x_1 and x_2 are completely co-dependent

assuming for simplicity: $\Sigma_{12} = \Sigma_{22} = \Sigma_{11} \Longrightarrow$

- $\mu_{x_1|x_2} = x_2 + (\mu_1 - \mu_2)$
- $\Sigma_{x_1|x_2} = 0$ (no varience!)

(3) GENERAL CASE:

Notice how $\Sigma_{12}\Sigma_{22}^{-1}$ adjusts <u>the mean</u> and $\underbrace{\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}}$ adjusts the varience based on <u>the covariatio</u>

how similar the covarience $(\Sigma_{21}, \Sigma_{12})$ is to the variations (Σ_{22}) in x_2.